
УДК 004.942:343.92

М. М. ЗАЦЕРКЛЯНИЙ,

доктор технічних наук,

професор кафедри інформаційних систем і технологій у діяльності ОВС

Харківський національний університет внутрішніх справ

А. С. БАБІЙ,

викладач кафедри інформаційних систем і технологій у діяльності ОВС

Харківський національний університет внутрішніх справ

ПОПЕРЕДНІЙ АНАЛІЗ ДАНИХ У СИСТЕМАХ ОБРОБКИ ІНФОРМАЦІЇ ПРО СКОЄНІ ЗЛОЧИНИ

На сьогодні опрацювання та впорядкування інформаційних потоків є однією з актуальних проблем управління. В цьому не є винятком і правоохоронні органи.

Інформація, зафіксована на основі офіційного обліку або шляхом спеціально організованого дослідження у статистичних картках, журналах обліку, базах даних, комп'ютерних інформаційних системах, в анкетах опитування громадян, при вивченні кримінальних, адміністративних, цивільних справ та інших матеріалів, є розрізненими «горами даних» про одиниці досліджуваної сукупності. Ця інформація стає дієвим засобом у боротьбі зі злочинністю тільки за умови її аналітичного опрацювання.

Бурхливі зміни в економічній сфері суспільного життя впливають на криміногенну ситуацію в державі. Деякі питання аналізу злочинності, а також створення інформаційних систем для обробки інформації про скоєні злочини розглядаються у роботах [1 – 6]. Проте існуючі методики аналізу інформації, що базуються на використанні систем збереження й пошуку кримінологічних даних та персоналу для аналізу масивів одержаної інформації є недостатньо ефективними.

На сьогодні існує потреба в системі підтримки прийняття рішень, яка б дозволяла комплексно оцінювати злочинність, адже завдання щодо вироблення рішень в органах внутрішніх справ доводиться

виконувати з огляду на обмеження, що накладає предметна сфера, а саме:

- неможливість повторення статистичних вимірів;
- короткі динамічні ряди статистичних

даних;

- врахування експертних оцінок;
- потреба в багатофакторних дослідженнях.



Рис. 1. Схема початкового аналізу динамічного ряду злочинності

Завдання, що виконуються при початковому аналізі динамічних рядів злочинності (рис. 1), як мають самостійне значення, так і входять до інших завдань. Зокрема, ці завдання виконуються перед побудовою статистичної моделі злочинності та її прогнозуванням із метою виявлення додаткової інформації, що може вплинути як на побудову моделі, так і на прогнозування та виявлення чинників, які істотно впливають на рівень злочинності в конкретному регіоні в конкретний проміжок часу.

Попередній аналіз даних. Під аномальним розуміють окреме значення рівня часового ряду, яке не відповідає потенційним можливостям досліджуваної системи і яке, залишаючись рівнем ряду, істотно впливає на значення основних характеристик часового ряду.

Для виявлення аномальних рівнів у динамічних рядах злочинності рекомендується використовувати метод Ірвіна. У цьому методі розрахункові значення критерію Ірвіна визначаються з використанням формул:

$$\lambda_i = \frac{|x_i - x_{i-1}|}{\sigma_x}; i=2, \dots, n,$$

де σ_x – середнє квадратичне відхилення, яке розраховується так:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}; \bar{x} = \frac{\sum_{i=1}^n x_i}{n}.$$

Розрахункові значення $\lambda_2, \lambda_3, \dots, \lambda_n$ порівнюються з табличними значеннями критерію Ірвіна λ_α і, якщо виявляються більшими за табличні, то відповідне значення x_i рівня ряду слід вважати аномальним.

Так, наприклад, за наявності об'єктивних причин для появи аномального рівня динамічного ряду в певний період у межах України в деякому регіоні аномальний рівень відсутній. Це може, зокрема, говорити про незадовільну роботу відповідного підрозділу МВС із реєстрації злочинів та правопорушень. Наявність аномального рівня в певному регіоні при його відсутності в інших свідчить про певні особливості цього регіону (якість роботи правоохоронних органів, специфіка нормативних актів органів місцевого самоврядування тощо).

Рішення про врахування чи неврахування аномальних рівнів у дослідженні приймає користувач системи, експерт. Програмна система звертає увагу користувача на аномальний рівень та, оцінюючи значення аномального рівня відносно попередніх та наступних рівнів ряду, а також значень інших доступних рядів злочинності (в інших регіонах та в цілому в Україні), пропонує користувачу прийняти відповідне рішення.

Для розв'язування задачі виявлення трендової складової пропонується метод, який можна подати у вигляді чотирьох етапів.

1. Початковий часовий ряд $(x_1, x_2, x_3, \dots, x_n)$ розбивається на дві приблизно однакові за кількістю рівнів частини: у першій частині n_1 перших рівнів ряду, у другій – n_2 інших рівнів ($n_1 + n_2 = n$).

2. Для кожної з цих частин обчислюються середні значення і дисперсії:

$$\bar{x}_1 = \frac{\sum_{t=1}^{n_1} x_t}{n_1}; \sigma_1^2 = \frac{\sum_{t=1}^{n_1} (x_t - \bar{x}_1)^2}{n_1 - 1};$$

$$\bar{x}_2 = \frac{\sum_{t=n_1+1}^n x_t}{n_2}; \sigma_2^2 = \frac{\sum_{t=n_1+1}^n (x_t - \bar{x}_2)^2}{n_2 - 1}.$$

3. З'ясовується однорідність дисперсій обох частин ряду за допомогою F -критерію Фішера, який ґрунтується на порівнянні розрахункового значення цього критерію

$$F = \begin{cases} \frac{\sigma_1^2}{\sigma_2^2}, & \text{якщо } \sigma_1^2 > \sigma_2^2, \\ \frac{\sigma_2^2}{\sigma_1^2}, & \text{якщо } \sigma_1^2 < \sigma_2^2 \end{cases}$$

з критичним значенням F_α , знайденим у відповідній таблиці залежно від заданого рівня значення α .

Якщо розрахункове значення F менше за табличне F_α , то гіпотеза про однорідність дисперсій приймається. Якщо значення F більше або дорівнює F_α , гіпотеза про однорідність дисперсій відхиляється і робиться висновок, що даний метод не дає відповіді на наявність чи відсутність тренда.

4. Перевіряється гіпотеза про відсутність тренда з використанням t -критерію Стьюдента. Для цього знаходиться розрахункове значення цього критерію за формулою:

$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}},$$

де σ – середнє квадратичне відхилення різниці середніх:

$$\sigma = \sqrt{\frac{(n_1 - 1)\sigma_1^2 + (n_2 - 1)\sigma_2^2}{n_1 + n_2 - 2}}.$$

Якщо розрахункове значення t менше за табличне значення статистики Стьюдента t_α , то гіпотеза приймається, тобто тренда немає, якщо навпаки – тренд є.

Зауважимо, що в даному випадку табличне значення $t\alpha$ береться залежно від заданого рівня значення α і числа ступенів свободи, що дорівнює $n_1 + n_2 - 2$.

Так, зокрема, неузгодженість поведінки тренда в певному регіоні з тенденцією зміни злочинності в інших регіонах може говорити як про хорошу, так і про погану роботу відповідного підрозділу органу внутрішніх справ чи про певні особливості регіону.

Має самостійне значення і задача «Оцінювання тренд-сезонної адитивної моделі», оскільки сезонність пояснюється

не тільки зміною природно-кліматичних умов, а й іншими причинами як об'єктивного, так і суб'єктивного характеру. Дослідження сезонності розглянуто в роботі [6].

Висновки. Розв'язування задач, які є складниками попереднього аналізу динамічного ряду, потребує залучення та врахування додаткової інформації, а саме експертних оцінок, результатів аналізу діяльності правоохоронних органів чи соціально-економічного розвитку конкретного регіону.

Література

1. Кудрявцев В. Н. Основные причины организованной преступности в России // Вестник Российской академии наук. – 1999. – № 9. – Т. 69. – С. 794.
2. Бандурка А. М. Преступность в Украине: причины и противодействие : моногр. / А. М. Бандурка, Л. М. Давыденко. – Х., 2003. – С. 32.
3. Вицин С. Е. Системный подход в статистике / Вицин С. Е. – М., 1989.
4. Социально-экономическая статистика : учеб. для вузов / [под ред. проф. Б. И. Башкатова]. – М. : ЮНИТИ-ДАНА, 2002. – 703 с.
5. Савюк Л. К. Правовая статистика / Савюк Л. К. – М., 2001.
6. Зацеркляний М. М. Автоматизація аналізу сезонних коливань рівня злочинності / М. М. Зацеркляний, А. С. Бабій // Право і безпека. – 2005. – № 4'3.

Надійшла до редколегії 21.04.2009

Анотації

У статті розглядаються підходи для створення інформаційної системи, призначеної для моделювання злочинності. В роботі пропонується підсистема попереднього аналізу даних для обробки інформації про скоєні злочини. Розглядаються етапи та методики первинної обробки даних.

В статье рассматриваются подходы для создания информационной системы, предназначенной для моделирования преступности. В работе предложена подсистема предварительного анализа данных для обработки информации о совершенных преступлениях. Рассматриваются этапы и методики первичной обработки данных.

In the article approaches are considered for creation of the informative system, intended for the design of criminality. The subsystem of preliminary analysis of data is in-process offered for treatment of information about perfect crimes. The stages and methods of roughing-out of information are examined.